

Оценка эффективности обучения сверточных нейронных сетей на синтетических данных для распознавания символов государственных регистрационных знаков транспортных средств

Галимуллин Дамир Зиннурович

Казанский инновационный университет

доцент кафедры информационных технологий и безопасности

Денисов Алексей Игоревич

ООО «Вероятность»

директор

Илларионов Сергей Юрьевич

Казанский инновационный университет

студент кафедры информационных технологий и безопасности

Аннотация

В данной статье разработан и реализован программно механизм генерации синтетических обучающих выборок, имитирующих номерные знаки РФ. Создана собственная база размеченных данных, релевантных реальным условиям съема и возникающим искажениям. Изучены базовые архитектуры нейронных сетей глубокого обучения. Предложены конфигурации сетей для работы с синтетическими данными, реализован механизм обучения на основе искусственной обучающей выборки, произведена оценка по распознаванию символов номерных знаков.

Ключевые слова: распознавание изображений, машинное обучение, нейронные сети глубокого обучения, сверточные нейронные сети, разметка данных, синтетические обучающие выборки.

Estimation of efficiency of training of convolutional neural networks on synthetic data for recognizing symbols of state registration marks of vehicles

Galimullin Damir Zinnyrovich

Kazan Innovation University

Docent of the Department of Information Technology and Security

Denisov Alexey Igorevich

ООО «Veroyatnost»

Director

Illarionov Sergey Yurievich

Kazan Innovation University

Student of the Department of Information Technology and Security

Abstract

In this article, a mechanism for generating synthetic training samples simulating the license plates of the Russian Federation has been developed and implemented programmatically. Created their own database of marked data, relevant to the real conditions of removal and the resulting distortions. The basic architectures of neural networks of deep learning are studied. Configurations of networks for working with synthetic data are proposed, a learning mechanism based on an artificial learning sample is implemented, an evaluation is made for the recognition of symbols.

Keywords: Image recognition, machine learning, neural networks of deep learning, convolutional neural networks, data markup, synthetic training samples

ВВЕДЕНИЕ

В настоящее время на рынке систем видеофиксации имеется ряд продуктов, позволяющих обеспечить высокую вероятность распознавания номеров транспортных средств. Ключевым элементом таких систем является программное обеспечение, позволяющее распознавать номера.

В период с 2010 года по настоящее время в мире получила широкое развитие новая технология распознавания изображений, основанная на применении нейронных сетей глубокого обучения [1-3]. В настоящее время данная технология достигла высокой степени зрелости и доступна для коммерческого использования в режиме открытого исходного кода. Применение данной технологии для распознавания номеров транспортных средств, как ожидается, позволит достичь высокой вероятности распознавания при использовании относительно недорогой программной и аппаратной части. Более того, данная технология имеет высокий потенциал для распознавания номеров в сложных условиях (плохая видимость, геометрические искажения, наложение шума разного характера и т.д.). Результаты использования данной технологии в других областях показывают, что вероятность распознавания изображений с помощью глубокого обучения оказывается достаточно высокой. Также за счет применения данных технологий при разработке программной части позволяет существенно снизить стоимость системы видеофиксации высокого качества, что позволяет расширить область применения и обеспечивает высокий коммерческий потенциал внедрения такого программного продукта.

Первые успешные применения нейронных сетей глубокого обучения относятся к началу 90-х годов прошлого века. В AT&T Bell Labs под руководством Яна Лекуна (Yann LeCun) была разработана система автоматического распознавания рукописных цифр [4]. Однако в связи со сложностью обучения сверточных нейронных сетей в течение продолжительного периода времени интерес исследователей был сконцентрирован на других подходах к машинному обучению. Прогресс в аппаратной и научной области привел к возобновлению интереса научного сообщества к данному методу. Особенно сильный интерес появился после

2010-х годов, когда такие сети начали показывать лучшие результаты при распознавании изображений, речи и т.д. Системы, основанные на применении нейронных сетей глубокого обучения, в настоящее время показывают лучшие результаты в распознавании изображений на тестовых базах изображений, таких как MNIST и CIFAR-10. Дополнительным фактором, стимулирующим применение данного инструмента в различных областях обработки данных является прогресс аппаратных вычислительных мощностей. В частности, графические процессоры, при использовании соответствующего программного обеспечения, позволяют сократить время обучения в десятки раз. Появление различных библиотек с открытым исходным кодом существенно упростило практическое применение, и в настоящее время можно говорить о появлении технологии, доступной для широкого коммерческого и научного применения.

Необходимо сказать о главных целях проводимых работ, мотивации для их проведения. Существующие базы номерных знаков, используемые для обучения, достаточно ограничены как с точки зрения количества, так и качества. В свою очередь, современные нейронные сети весьма критичны к объему данных и их качеству. В этом случае, предпочтительным подходом представляется создание собственной базы данных, объем которой и релевантность данных реальным условиям съема и возникающим искажениям можно регулировать. Для решения этой задачи был разработан механизм генерации синтетических обучающих выборок. Далее было необходимо рассмотреть базовые архитектуры нейронных сетей глубокого обучения, конвертацию данных в рабочий формат сетей, построить функционирующие конфигурации для работы с синтетическими данными, реализовать механизм обучения на основе искусственной обучающей выборки, оценить эффективность технологии распознавания на искусственных данных. Эти процедуры рассмотрены ниже в статье.

ПОДГОТОВКА НАБОРА РАЗМЕЧЕННЫХ ДАННЫХ

Для обучения нейронной сети нам было необходимо подготовить размеченную базу данных. Нейронные сети, которые подробно рассматриваются в данной работе, являются одной из разновидностей алгоритмов машинного обучения, или machine learning. Машинное обучение – это один из подразделов искусственного интеллекта. Основным свойством алгоритмов machine learning является их способность обучаться в процессе работы. Например, не имея никакой предварительной информации о том, что представляют собой данные и какие в них существуют закономерности, а только некоторый входной набор объектов и значения некоторых признаков для каждого из них вместе с меткой класса, в процессе работы алгоритм сам выявляет скрытые закономерности, то есть обучается, и после обучения способен предсказывать класс уже для новых объектов, которые он не видел ранее [5].

Выделяют два основных типа машинного обучения: обучение с учителем и обучение без учителя. Обучение с учителем предполагает, что алгоритму, кроме самих исходных данных, предоставляется некоторая дополнительная информация о них, которую он может в дальнейшем использовать для обучения. Например, задачу классификации можно сформулировать следующим образом: имея некоторый набор объектов, каждый из которых относится к одному из нескольких классов, необходимо определить, к какому из этих классов относится новый объект. Задача распознавания номерных знаков является типичной разновидностью задачи классификации: есть разные символы на номере - классы, и задача алгоритма – «распознать» символ, то есть отнести его к одному из существующих классов.

Таким образом, нейронные сети как алгоритм машинного обучения, требуют обучения. Но в отличие от большинства алгоритмов нейронные сети очень критичны к объему данных, к объему той обучающей выборки, которая необходима для того, чтобы их натренировать. Известно, что на маленьком либо некачественном объеме данных сети плохо работают. Они неудовлетворительно обобщают данные, плохо работают на примерах, которые они не видели в процессе обучения.

Задача подготовки качественной обучающей выборки являлась одной из важнейших задач, причем, было необходимо, чтобы подготовленные данные были релевантны, соответствовали реальным данным, получаемым при обработке видеопотока. Для этого нами был проведен детальный анализ реальной базы номеров (Рис.1) и были выделены следующие искажения, возникающие в кадре.

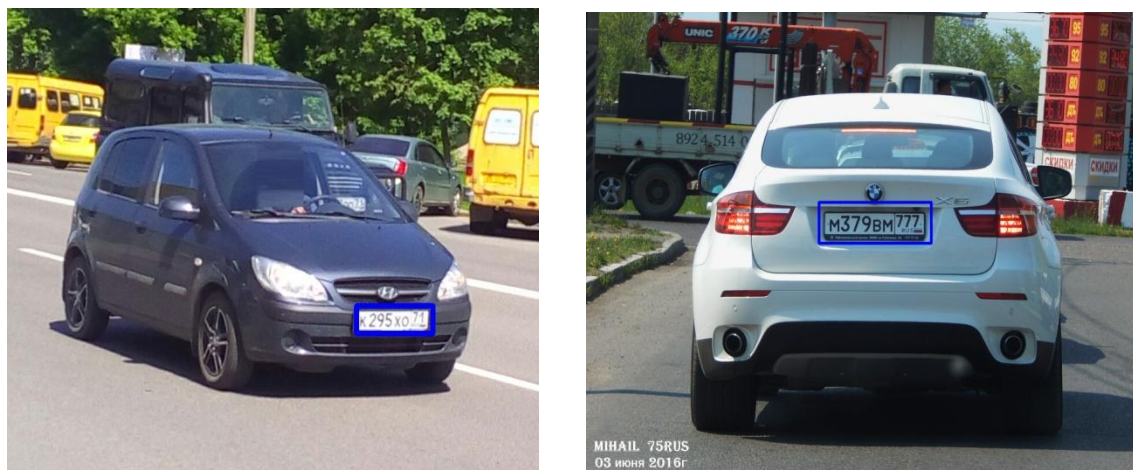


Рисунок 1 - Выделение области номера ТС при помощи каскадов Хаара

Отметим их:

1. Искажения самих символов, вызванные их «размытием» и обусловленные движением транспортного средства, погодными условиями, значительным расстоянием до камеры при фиксации интересующего объекта. Данный тип искажений представлен на рисунке 2.



Рисунок 2 - Искажения самих символов, возникающие на номере ТС

2. Искажения фона, вызванные шумом вследствие погодных условий, времени съема, качества камеры, либо загрязнением области номерного знака. Данный тип искажений представлен на рисунке 3.



Рисунок 3 - Искажения фона, возникающие на номере ТС

3. Искажения символов, вызванные из-за метода либо угла фиксации видеокамеры, формы транспортного средства и способа крепления номера. Такие искажения приводят к геометрическим искажениям формы символа либо изменению его наклона по сравнению со стандартной. Данный тип искажений представлен на рисунке 4.



Рисунок 4 - Искажения геометрической формы символов в виде наклона вправо, возникающие на номере ТС

Поскольку создание реальной базы размеченных номеров, где включались бы все возможные искажения, достаточно затруднительно вследствие ограниченности доступа к такого рода данным, то, после проведенного анализа искажений, было решено:

1) Создать механизм генерации искусственных номерных знаков на основе реального прототипа

2) Подобрать преобразования, позволяющие смоделировать искажения, возникающие на номерных знаках, которые указаны в пунктах 1-3.

3) Выделить символы и цифры на модельном номерном знаке, используя окно с заранее определенными и фиксированными размерами

4) Занести выделенные символы и цифры в базу данных с указанием их класса для формирования обучающей выборки. База данных должна была иметь следующий вид (см. табл.1)

Таблица 1. - Структура базы данных

Символ или цифра	Метка класса	Символ или цифра	Метка класса	Символ или цифра	Метка класса
1	1	А	11	Т	21
2	2	В	12	Х	22
3	3	С	13	У	23
4	4	Д	14	Шум	0
5	5	Е	15		
6	6	Н	16		
7	7	К	17		
8	8	М	18		
9	9	О	19		
10	10	Р	20		

Класс с меткой «0» предполагает шумовую выборку, составленную из фрагментов изображений номера ТС, которые не включают символы или цифры полностью, либо содержат только фон.

Отметим преимущества синтетической обучающей выборки:

- 1) Возможность сформировать выборку практически неограниченной размерности, что позволяет проводить эффективно обучение
- 2) Смоделировать те искажения, которые редко попадают в реальной выборке вследствие ее ограниченности
- 3) Возможность сформировать смешанную выборку, включающую в себя как искусственные, так и реальные данные.

Для создания искусственных номерных знаков и моделирования искажений нами был проведен ряд преобразований. Рассмотрим их подробно.

1. Брался эталонный номер согласно ГОСТ Р 50577-93 «Знаки государственные регистрационные транспортных средств. Типы и основные размеры. Технические требования» [6]. Учитывая, что большинство регистрационных знаков имеют вид, указанный на рисунке ниже, он являлся шаблоном для последующей генерации искусственных номерных знаков (Рисунок 5)



Рисунок 5 - Эталонный номер

2. На исходном эталонном номерном знаке после бинаризации рассчитывались области, где расположены символы (Рисунок 6)

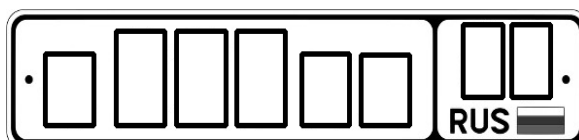


Рисунок 6 - Определение областей расположения символов

3. Начертание символов определяется шрифтом ЖР5 по ГОСТ 3489.2-71 «Шрифты типографские. Гарнитура Журнальная рубленая» [7]. На рисунке ниже приведена область из ГОСТа начертания шрифта. Данный эталонный шаблон использовался для выделения необходимых символов

1234567890
АВСДЕКМОРТХУ

Рисунок 7 - Начертание символов шрифта для автомобильных знаков

4. При помощи генератора случайных чисел вырезались символы из области шаблона, представленного на рисунке 7. Далее они помещались в области, представленные на рисунке 6, согласно их расположению по ГОСТ Р 50577-93. Таким образом, получался синтезированный номерной знак (Рисунок 8)



Рисунок 8 - Искусственный номерной знак

5. На генерируемые номерные знаки наносился шумовой фон и формировался эффект «теней» (Рисунок 9). Шумовое распределение моделировалось распределением Гаусса, эффект «теней» при помощи равномерного распределения.



Рисунок 9 - Моделирование фона номерного знака

6. Далее, при помощи размывания во взаимно перпендикулярных направлениях при помощи масок Гаусса моделировалось «размытие» символов номерного знака



Рисунок 10 - Размывание символов номерного знака

7. Для того, чтобы изменить наклон букв и изменить их положение вследствие геометрических преобразований были использованы аффинное преобразование и преобразование перспективы. На рисунке 11 представлены примеры итоговых синтетических номеров. Аффинное преобразование осуществлялось на основе трех точек, которые образовывали прямоугольный треугольник. Случайным образом положения этих точек менялись в узком диапазоне. Преобразование перспективы использовало 4 точки, образующие квадрат, положения этих точек также менялись в узком диапазоне порядка 3-4 пикселей по оси X и Y.



Рисунок 11 - Генерируемые номера

8. Следующим этапом при помощи специально расположенного окна вырезались фрагменты, содержащие символы с нанесенными случайными искажениями. Далее, данные изображения заносились в базу с соответствующими метками

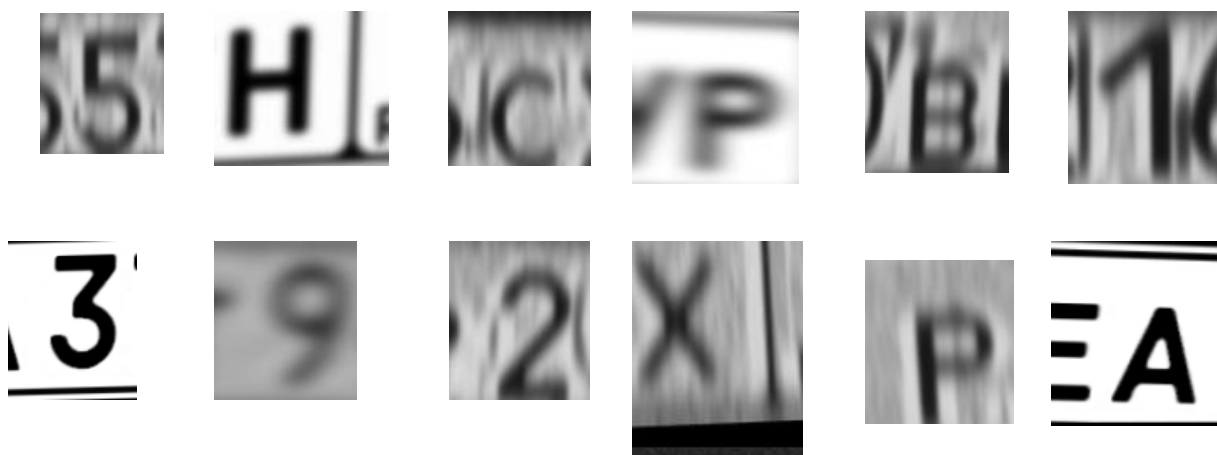


Рисунок 12 - Символы, генерируемые для базы

9. На следующем этапе было необходимо сформировать шумовую выборку. Шумовая выборка состояла из фрагментов номерных знаков без символов, либо с обрезанными фрагментами символов, либо из изображений, отображающих шум с нормальным распределением (Рисунок 13).

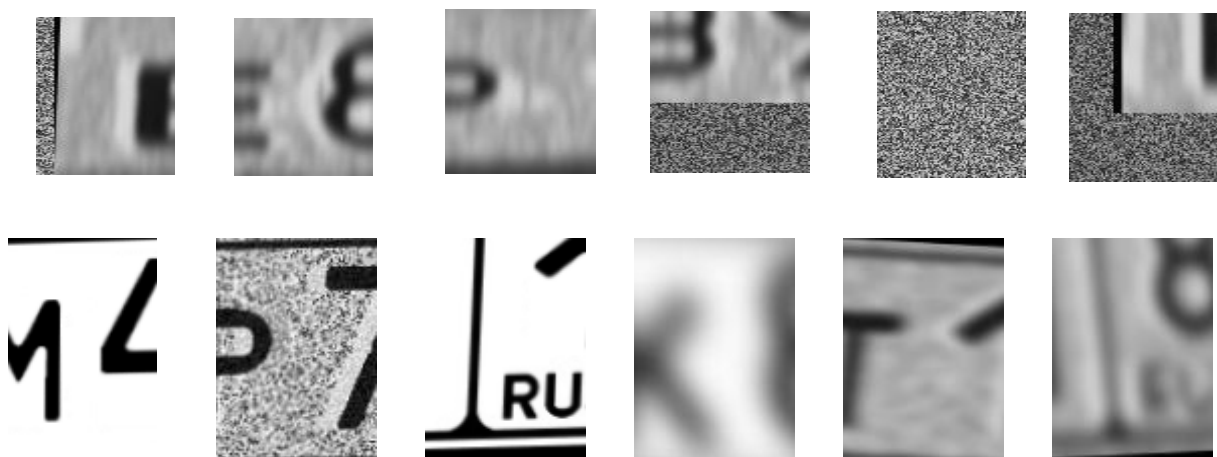


Рисунок 13 - Шумовые фрагменты, генерируемые для базы

Были созданы программы, позволяющие автоматизировать процесс генерации данных и записи в соответствующие директории.

Таким образом, на данном этапе была рассмотрена общая процедура создания набора размеченных данных. Несомненным достоинством данной процедуры является возможность создания базы данных необходимых размеров с различными типами искажений в отличие от ограниченных размеченных данных по российским номерным знакам, присутствующих в свободном доступе, что позволяет эффективно работать с нейронными сетями.

ОБУЧЕНИЕ, ТЕСТИРОВАНИЕ И ВЫБОР НЕОБХОДИМОЙ КОНФИГУРАЦИИ

Для обучения были построены 3 конфигурации сетей на основе базовых архитектур Lenet и Cifar [8,9]. Конфигурация Lenet отлично зарекомендовала себя в работе с распознаванием рукописных символов в датасете MNIST, показывая на нем точность 99%, и соответственно, подходила для решения нашей задачи [10]. Конфигурация типа Cifar также представляла определенный интерес, так как распознавание номеров домов (цифр) на датасете SVHN с использованием данной конфигурации является классическим, достигая точности 96%. Таким образом, эти конфигурации представляли для нас практический интерес. В основе данных архитектур лежит совокупность сверточных слоев, слоев пространственного объединения и полносвязных слоев. Использовались так называемые конфигурации Cifar-full и Cifar-quick [9], которые мы будем называть сокращенно Cifar-f и Cifar-q, с некоторыми модификациями слоев.

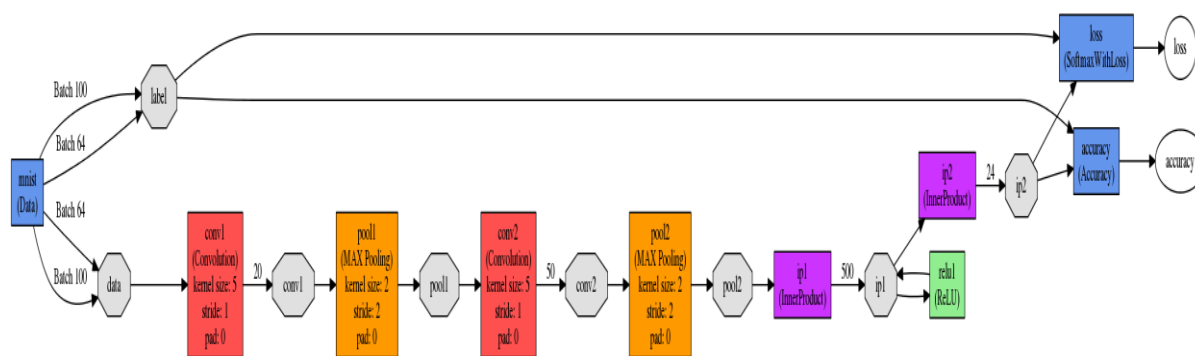


Рисунок 14 - Конфигурация Lenet сверточной нейронной сети

Конфигурация Lenet представлена 2 сверточными слоями (conv1, conv2), 2 слоями пространственного объединения (pool1, pool2) и двумя полносвязными слоями (ip1, ip2). Она представлена на рисунке 14.

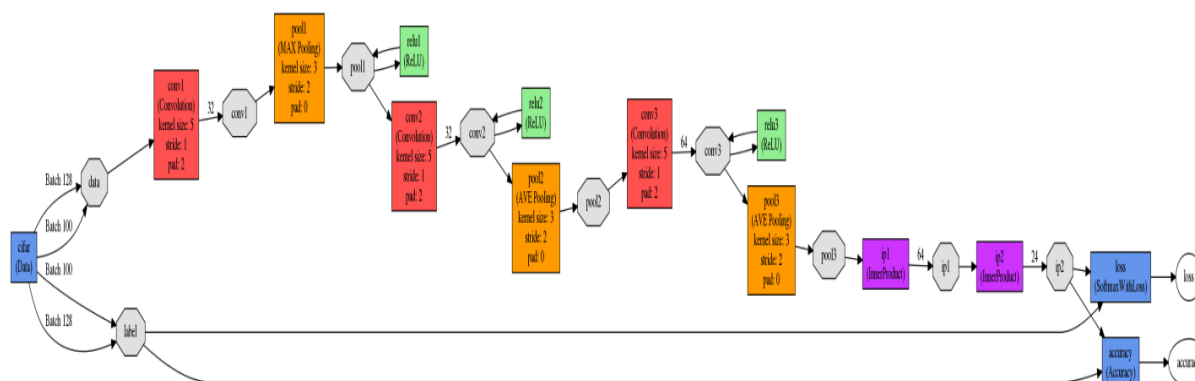


Рисунок 15 - Конфигурация Cifar-q сверточной нейронной сети

Конфигурация Cifar-q представлена 3 сверточными слоями (conv1, conv2, conv3), 3 слоями пространственного объединения (pool1, pool2, pool3) и двумя полносвязными слоями (ip1, ip2). Она представлена на рисунке 15. Фактически, данная конфигурация повторяет структуру сети Lenet, усложняя ее путем добавления добавочной совокупности из одного сверточного слоя и слоя пространственного объединения.

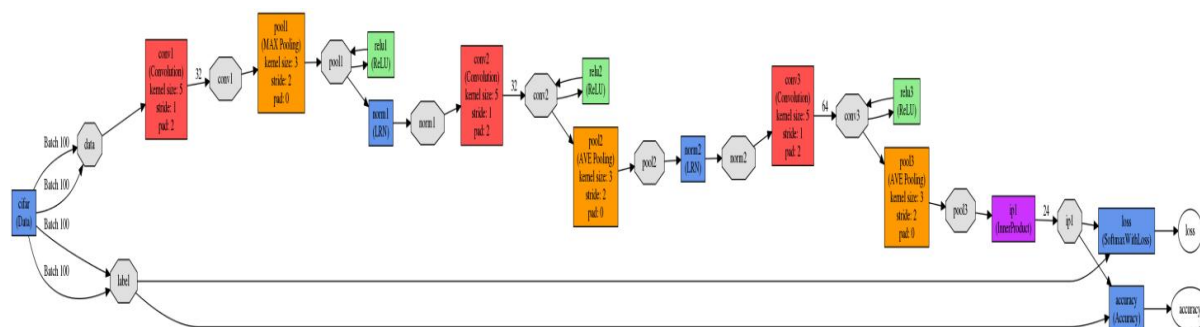


Рисунок 16 - Конфигурация Cifar-f сверточной нейронной сети

Конфигурация Cifar-f представлена 3 сверточными слоями (conv1, conv2, conv3), 3 слоями пространственного объединения (pool1, pool2, pool3) с последующей нормализацией и двумя полносвязными слоями (ip1, ip2). Она представлена на рисунке 16.

Все сети обучались на одной обучающей выборке. Размер выборки составлял величину порядка 24000 изображений, в среднем, по 1000 изображений на каждый класс. Количество итераций составило 2500 для каждой конфигурации. Определенные параметры сетей подбирались эмпирически, путем перебора. Основным параметром, который обеспечивал сходимость всех конфигураций, являлась скорость обучения сети *base_lr*, определяемая в конфигурационном файле. Слишком высокое значение скорости приводило к прекращению обучения сети либо ее параличу. В итоге, мы последовательно уменьшали эту величину, чтобы обеспечить сходимость. В схеме Lenet использовалась схема регулирования скорости обучения по формуле:

$$lr = base_lr * (1 + gamma^{iter})^{-power}, \tag{1}$$

т.е. практически линейное уменьшение скорости с увеличением номера эпохи (Рисунок 17), что обеспечивало плавную сходимость.

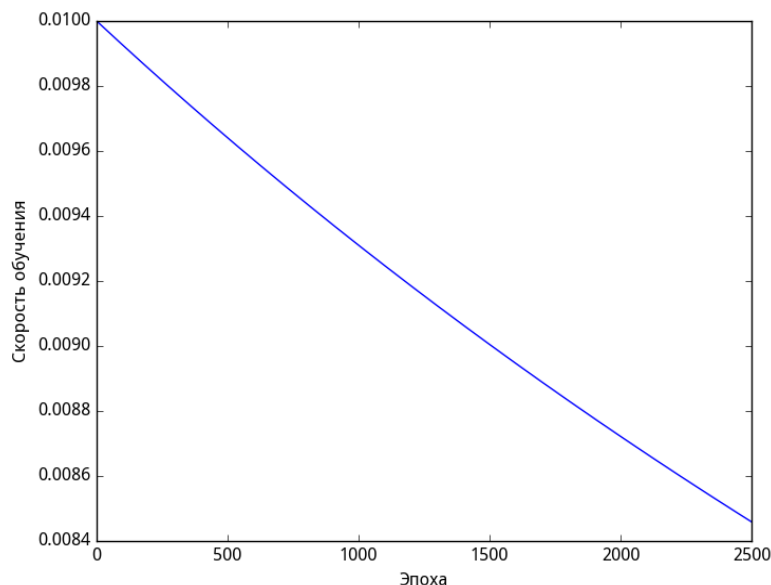


Рисунок 17 - Зависимость параметра скорости обучения от номера эпохи для конфигурации Lenet

Для конфигураций типа Cifar этот параметр являлся фиксированным, так как их скорость сходимости существенно меньше, чем у Lenet, и необходимость в регулировке не возникала.

Кривые, описывающие процесс обучения для всех конфигураций представлены ниже (Рисунок 18-20). Синяя линия показывает точность распознавания на обучающей выборке, зеленая и красная — процент ошибки на обучающем и валидационном датасете, соответственно. Из графиков на рисунках 18-20 видно, что обучение всех конфигураций прошло успешно: кривая точности, имеющая синий цвет, в пределе стремится к единице. Это означает, что данные из валидационного датасета распознаются успешно. Ошибки на обучающем и валидационном датасетах отображаются зеленой и красной кривыми и стремятся плавно к минимальному значению. Это означает, что как параметры сети, так и обучающие данные сформированы верно.

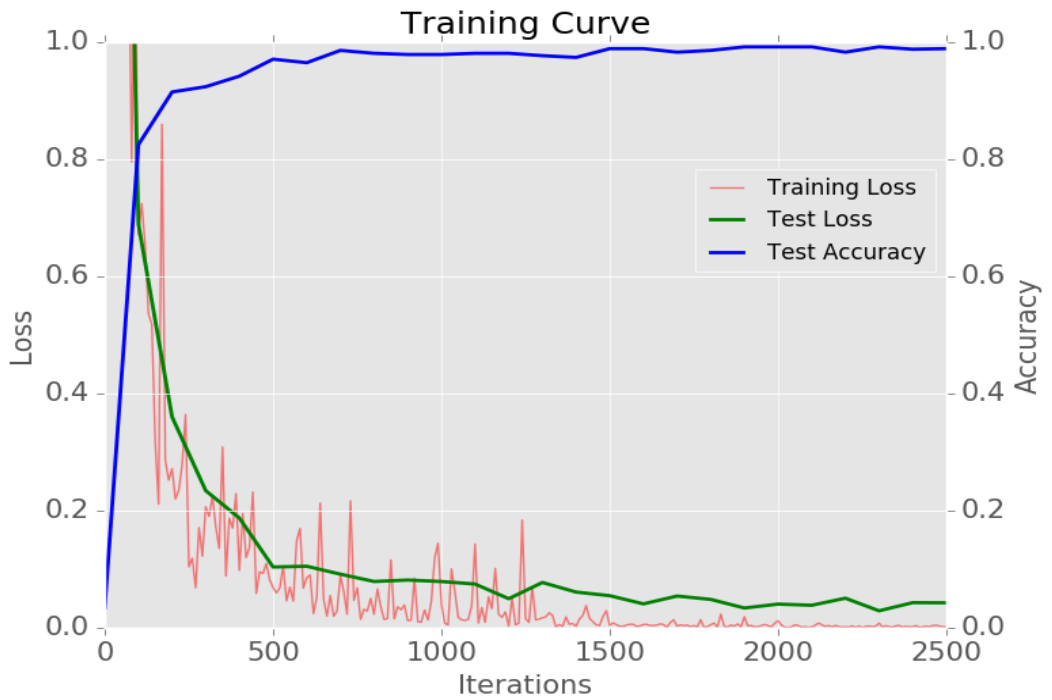


Рисунок 18 - Кривая обучения нейронной сети с конфигурацией LeNet



Рисунок 19 - Кривая обучения конфигурации Cifar-q нейронной сети



Рисунок 20 - Кривая обучения конфигурации Cifar-f сверточной нейронной сети

Точность во время обучения контролировалась на валидационном датасете, размер которого составлял 10% от обучающей выборки, и соответственно, не входил в нее. Точность конфигурации Lenet составила 98.9%, точность Cifar-f и Cifar-q — 98% и 97%, соответственно

Для независимой оценки качества построенной сети точность оценивалась на обучающем датасете. Загружались символы из обучающей выборки и определялась принадлежность к конкретному классу с оценкой ошибки в процентах, что иллюстрирует рисунок 21.

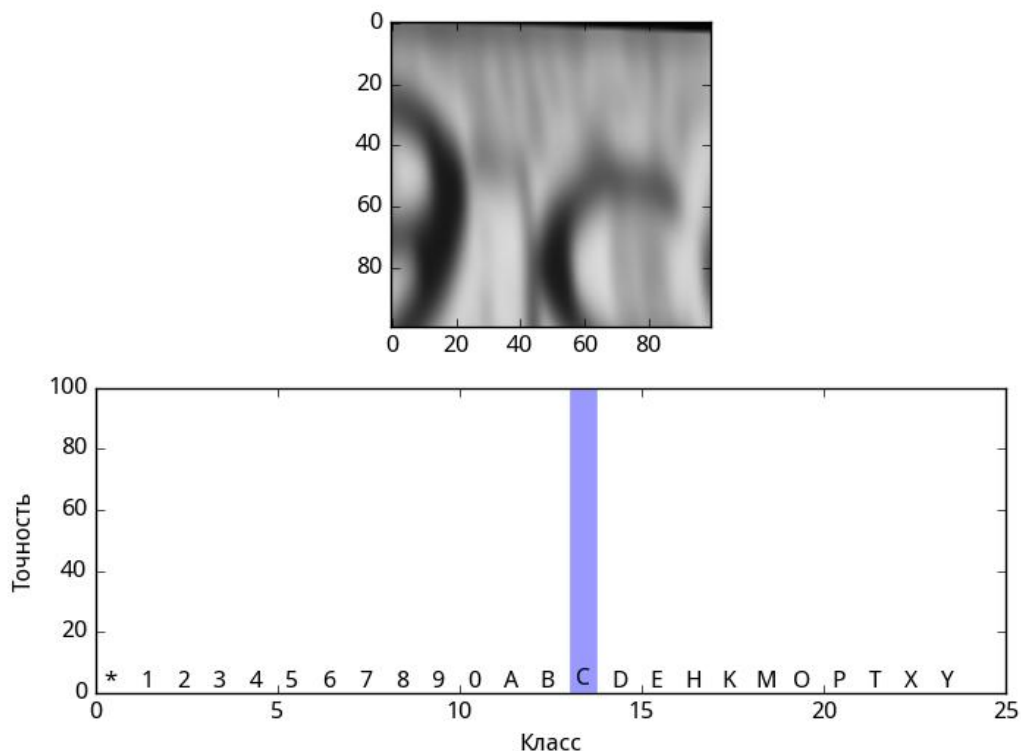


Рисунок 21 - Вероятность отнесения символа к соответствующему классу

Далее, через каждую сеть пропускался весь обучающий датасет, что позволило определить более точно ошибку по каждому классу, ее характер, и выбрать более предпочтительную конфигурацию. Результаты данной процедуры представлены на гистограммах из рисунков 22-24. Из них видно, что сеть типа Lenet работает на всех символах с ошибкой не более 1%, за исключением символа «O», что легко объяснимо, так как сходство символов «0» и «O» приводит к более значительной ошибке для этого символа. Результаты работы конфигураций класса Cifar хуже: ошибка для цифр достигала 1%, а для букв — 2% и более.

Таким образом, проведенный анализ гистограмм и полученные в результате обучения значения ошибок практически совпали и определили выбор сети Lenet для решения задачи распознавания символов и показали, что синтетические данные могут быть использованы для обучения сверточных нейросетей.

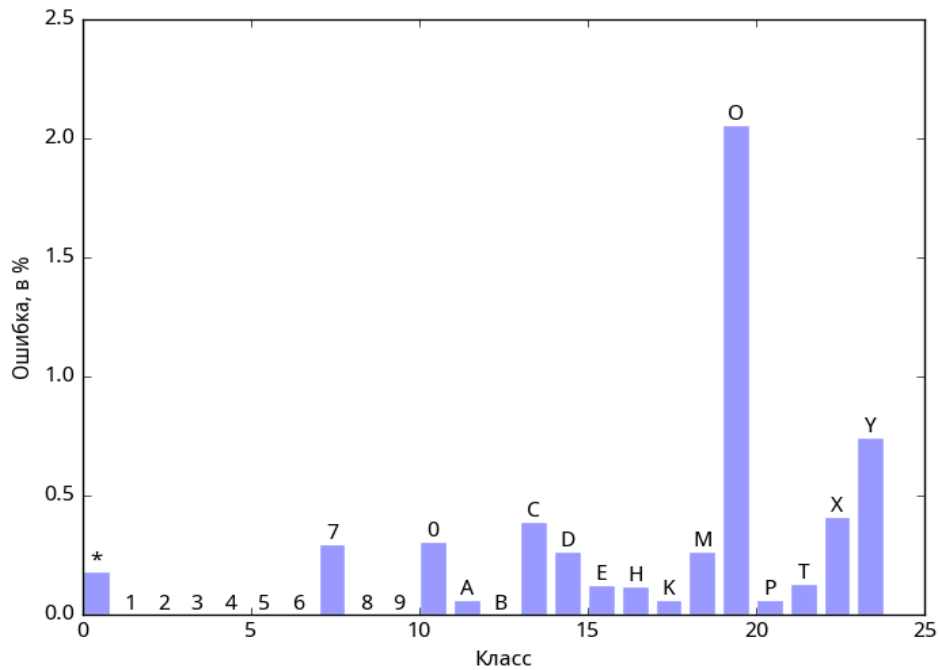


Рисунок 22 - Ошибка (в %) определения символа для конфигурации нейронной сети типа LeNet

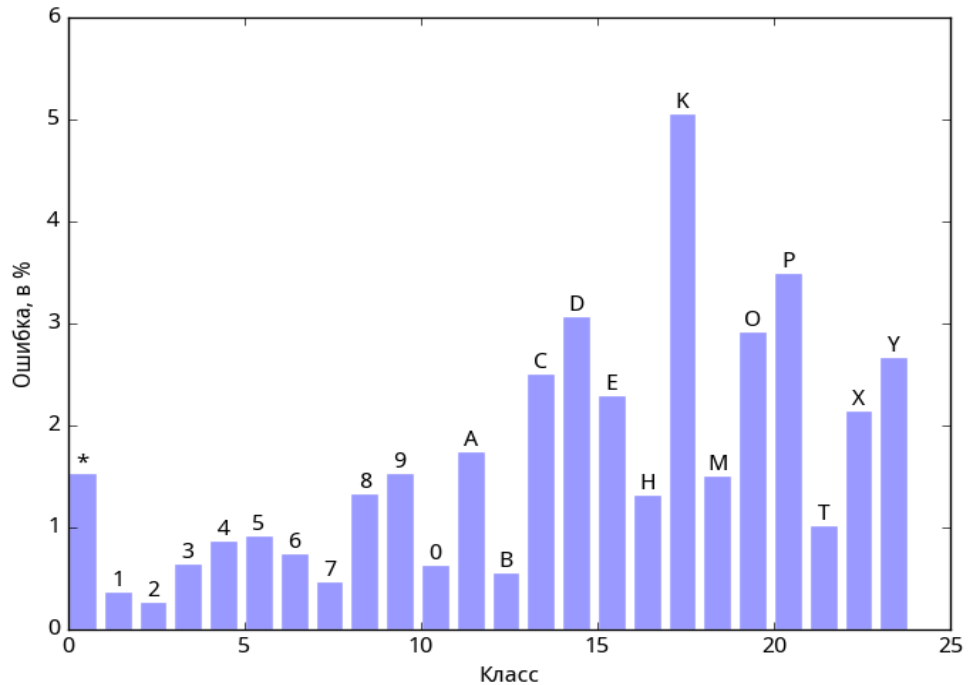


Рисунок 23 - Ошибка (в %) определения символа для конфигурации нейронной сети типа Cifar-q

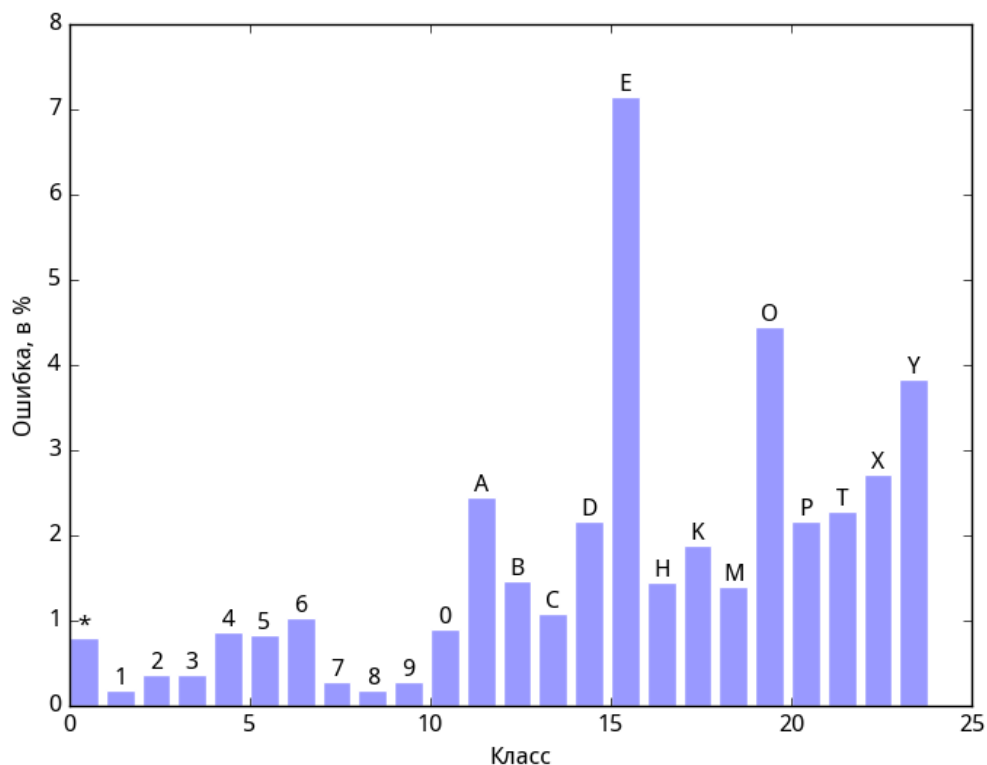


Рисунок 24 - Ошибка (в %) определения символа для конфигурации нейронной сети типа Cifar-f

ЗАКЛЮЧЕНИЕ

На основе анализа искажений, возникающих при видеофиксации номерных знаков, впервые было предложено построить синтетическую базу размеченных данных для обучения глубоких нейронных сетей. Был реализован алгоритм генерации искусственных номерных знаков с возможностью внесения различных искажений, имитирующих реальные. С использованием данного алгоритма была создана база размеченных данных, включающая в себя символы, определенные согласно ГОСТ Р 50577-93, и их метки.

Далее, на основе анализа ряда параметров и литературных источников, была выбрана и изучена библиотека глубоких нейронных сетей Caffe, проведена ее компиляция и установка. Для подготовки и конвертации размеченных данных в рабочий формат сети был создан ряд утилит.

После проведенного анализа литературных и сетевых источников были выбраны базовые архитектуры нейронных сетей глубокого обучения, предназначенные для решения задачи идентификации символов при различных искажениях. На их основе была осуществлена имплементация различных рабочих конфигураций с настройкой и подбором ряда параметров. Проведенное обучение показало высокую эффективность предложенных

конфигураций. Точность на валидационном датасете составила не менее 98% для каждой из них.

Проведенные расчеты по точности на каждом классе с использованием обучающего набора размеченных данных позволили выбрать базовую архитектуру LeNet, для которой точность составила не менее 99%. Таким образом, синтетические выборки могут быть использованы для обучения сверточных нейросетей, которые, в свою очередь, будут применяться для анализа реальных данных.

Библиографический список

1. Sermanet P., Chintala S. and LeCun Y. Convolutional Neural Networks Applied to House Numbers Digit Classification // International Conference on Pattern Recognition (ICPR 2012), 2012. С.3288-3291.
2. Dundar A., Bates J., Farabet C., Culurciello E. Tracking with deep neural networks // 47th Annual Conference CISS. 2013. С.1-5
3. Goodfellow I. J., Bulatov Y., Julian I., Arnaud S., Shet V. Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks // [Электронный ресурс]. Режим доступа: <http://static.googleusercontent.com/media/research.google.com/en//pubs/archive/42241>.
4. LeCun Y., Bottou L., Bengio Y. and Haffner P. Gradient-Based Learning Applied to Document Recognition // Proceedings of the IEEE. 1998. Т.86. №11. С.2278-2324.
5. Wu X., · Kumar V., · Ross Quinlan J, ·at all. Top 10 algorithms in data mining // Knowl. Inf. Syst. 2008. Т.14. С.1–37. DOI 10.1007/s10115-007-0114-2.
6. ГОСТ Р 50577-93. Знаки государственные регистрационные транспортных средств. Типы и основные размеры. Технические требования. Введен с 1993-06-29. – М.: ИПК Издательство стандартов, 2002. – 27 с.
7. ГОСТ 3489.2-71. Шрифты типографские. Гарнитура Журнальная рубленая (для алфавитов на русской и латинской графических основах). Назначение. Рисунок. Линия шрифта. Емкость. Введен с 01.01.1973. - 5 с. [Электронный ресурс]. Режим доступа: <http://www.internet-law.ru/gosts/gost/42532/>
8. Chaudhari M., Sondur S., Vanjare G. A review on Face Detection and study of Viola Jones method // International Journal of Computer Trends and Technology (IJCTT). 2015. Т.25. №1. С.54-61.
9. Caffe/models/ [Электронный ресурс]. Режим доступа: <https://github.com/BVLC/caffe/tree/master/model>
10. The MNIST database of handwritten digits. [Электронный ресурс]. Режим доступа: <http://yann.lecun.com/exdb/mnist/>